

RECEIVED

MAR 04 2002

TECH CENTER 1600/2900



OIPE

RAW SEQUENCE LISTING

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:28

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

RECEIVED

MAR 05 2002

TECH CENTER 1600/2900

SEQUENCE LISTING

1 (1) GENERAL INFORMATION:  
2 (i) APPLICANT: KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU  
3 KENKYUJO  
4 (ii) TITLE OF INVENTION: POLYPEPTIDE HAVING  $\alpha$ -FRUCTOFURANOSIDASE  
5 ACTIVITY  
6 (iii) NUMBER OF SEQUENCES: 6  
7 (iv) CORRESPONDENCE ADDRESS:  
8 (A) ADDRESSEE: BROWDY AND NEIMARK  
9 (B) STREET: 419 Seventh Street, N.W., Suite 300  
10 (C) CITY: Washington  
11 (D) STATE: D.C.  
12 (E) COUNTRY: USA  
13 (F) ZIP: 20004  
14 (v) COMPUTER READABLE FORM:  
15 (A) MEDIUM TYPE: Floppy disk  
16 (B) COMPUTER: IBM PC compatible  
17 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
18 (D) SOFTWARE: Patent In Release #1.0, Version #1.30  
19 (vi) CURRENT APPLICATION DATA:  
C--> 20 (A) APPLICATION NUMBER: US/09/986,682B  
C--> 21 (B) FILING DATE: 06-Nov-2001  
22 (C) CLASSIFICATION:  
23 (vii) PRIOR APPLICATION DATA:  
24 (A) APPLICATION NUMBER: JP 170,630/1996  
25 (B) FILING DATE: 10-JUN-1996  
26 (viii) ATTORNEY/AGENT INFORMATION:  
27 (A) NAME: BROWDY, Roger L.  
28 (B) REGISTRATION NUMBER: 25,618  
29 (C) REFERENCE/DOCKET NUMBER: TSUSAKI=2  
30 (ix) TELECOMMUNICATION INFORMATION:  
31 (A) TELEPHONE: 202-628-5197  
32 (B) TELEFAX: 202-737-3528  
33 (2) INFORMATION FOR SEQ ID NO: 1:  
34 (i) SEQUENCE CHARACTERISTICS:  
35 (A) LENGTH: 21 amino acids  
36 (B) TYPE: amino acid  
37 (D) TOPOLOGY: linear  
38 (v) FRAGMENT TYPE: N-terminal fragment  
39 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
40 Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr  
41 1 5 10 15  
42 Arg Ala Asp Met Leu

ENTERED

## RAW SEQUENCE LISTING

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:28

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

```

43          20
45 (2) INFORMATION FOR SEQ ID NO: 2:
46   (i) SEQUENCE CHARACTERISTICS:
47       (A) LENGTH: 12 amino acids
48       (B) TYPE: amino acid
49       (D) TOPOLOGY: linear
50   (v) FRAGMENT TYPE: internal fragment
51   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
52   Ser Val Phe Asp Gly Gly Asp Gly Thr Val Tyr Gln
53   1          5          10
55 (2) INFORMATION FOR SEQ ID NO: 3:
56   (i) SEQUENCE CHARACTERISTICS:
57       (A) LENGTH: 455 amino acids
58       (B) TYPE: amino acid
59       (D) TOPOLOGY: linear
60   (ii) MOLECULE TYPE: peptide
61   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
62   Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr
63   1          5          10          15
64   Arg Ala Asp Met Leu Lys Ile Pro Gly Gln Gln Asn Ser Pro Gln Phe
65   20          25          30
66   Lys Val Pro Gln Phe Asn Ala Ser Ala Ile Lys Asn Ile Asp Ser Ala
67   35          40          45
68   Lys Gly Tyr Asp Lys Ser Gly Asn Leu Ile Asp Leu Asp Val Trp Asp
69   50          55          60
70   Ser Trp Pro Leu Gln Asn Ala Asp Gly Thr Ala Ala Asn Tyr His Gly
71   65          70          75          80
72   Tyr His Ile Val Ser Ala Leu Ala Gly Asp Pro Lys Asn Ser Asp Asp
73   85          90          95
74   Thr Pro Leu His Leu Phe Tyr Gln Lys Val Gly Asp Thr Ser Ile Asp
75   100         105         110
76   Ser Trp Lys Asn Ala Gly Arg Val Phe Glu Asp Met Asp Lys Phe Val
77   115         120         125
78   Pro Asn Asp Pro Tyr Leu Lys Tyr Gln Thr Gln Glu Trp Ser Gly Ser
79   130         135         140
80   Ala Thr Leu Thr Lys Asp Gly Gln Val Arg Leu Phe Tyr Thr Asp Tyr
81   145         150         155         160
82   Ser Gly Asn Pro Glu Asp Gly Gly Thr Gly Ala Gly Asn Gln Ile Ile
83   165         170         175
84   Ser Thr Ala Gln Val Asn Leu Ser Gln Pro Asp Ala Ala Thr Leu Lys
85   180         185         190
86   Val Asp Gly Val Ser Asp His Lys Ser Val Phe Asp Gly Gly Asp Gly
87   195         200         205
88   Thr Val Tyr Gln Asn Ile Gln Gln Phe Ile Asp Glu Gly Lys Trp Ile
89   210         215         220
90   Ser Gly Asp Asn His Thr Leu Arg Asp Pro His Tyr Val Glu Asp Lys
91   225         230         235         240
92   Gly His Lys Tyr Leu Val Phe Glu Ala Asn Thr Gly Thr Thr Asp Gly
93   245         250         255

```

## RAW SEQUENCE LISTING

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:28

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

```

94   Tyr Gln Gly Asp Gln Ser Phe Asn Asn Lys Ala Tyr Tyr Gly Gly Ser
95               260                265                270
96   Asp Val Phe Phe Gln Asn Glu Lys Asn Lys Leu Leu Gln Ser Pro Lys
97               275                280                285
98   Lys Gln Ile Ala Ser Leu Ala Asn Gly Ala Leu Gly Ile Val Glu Leu
99               290                295                300
100  Ala Asp Asp Tyr Thr Val Lys Ser Val Met Lys Pro Leu Val Ala Ser
101               305                310                315                320
102  Asn Thr Val Ala Asp Glu Val Glu Arg Ala Asn Ile Phe Lys Met Asn
103               325                330                335
104  Asn Lys Trp Tyr Leu Phe Thr Asp Ser Arg Gly Ser Lys Met Thr Ser
105               340                345                350
106  Asp Gly Ile Asn Asp Lys Asp Val Tyr Met Leu Gly Pro Gly Gly Asp
107               355                360                365
108  Ser Leu Asn Gly Pro His Asn Pro Ile Asn Glu Thr Gly Leu Val Leu
109               370                375                380
110  Asn Met Asn Leu Asp Pro Ala Asp Leu Thr His Thr Tyr Ser His Cys
111               385                390                395                400
112  Gly Ile Pro His Pro Glu Gly Asn Asn Val Val Leu Thr Ser Tyr Met
113               405                410                415
114  Thr Asn Arg Gly Phe Tyr Pro Glu His His Ser His Leu Arg Asp Lys
115               420                425                430
116  Leu Gly Val Asn Ile Lys Gly Ser Asp Thr Ser Gly Gly Glu Asn Ser
117               435                440                445
118  Ser Gly Gln Gly Gln Phe Pro
119               450                455

```

## 121 (2) INFORMATION FOR SEQ ID NO: 4:

## 122 (i) SEQUENCE CHARACTERISTICS:

- 123 (A) LENGTH: 1365 base pairs
- 124 (B) TYPE: nucleic acid
- 125 (C) STRANDEDNESS: double
- 126 (D) TOPOLOGY: linear

## 127 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

128  ATG AAC AGC GGG GAC TAC AAG GAA GAC TAT GGT TTT GCC CAT ATT ACA      48
129  Met Asn Ser Gly Asp Tyr Lys Glu Asp Tyr Gly Phe Ala His Ile Thr
130  1      5      10      15
131  CGC GCT GAC ATG CTA AAA ATT CCA GGA CAA CAA AAC AGT CCT CAA TTT      96
132  Arg Ala Asp Met Leu Lys Ile Pro Gly Gln Gln Asn Ser Pro Gln Phe
133  20      25      30
134  AAA GTG CCT CAA TTC AAT GCA TCA GCA ATC AAA AAC ATT GAT TCG GCA      144
135  Lys Val Pro Gln Phe Asn Ala Ser Ala Ile Lys Asn Ile Asp Ser Ala
136  35      40      45
137  AAA GGG TAT GAT AAG TCA GGC AAC TTA ATA GAT TTA GAT GTA TGG GAT      192
138  Lys Gly Tyr Asp Lys Ser Gly Asn Leu Ile Asp Leu Asp Val Trp Asp
139  50      55      60
140  AGC TGG CCA CTG CAA AAC GCT GAT GGT ACT GCG GCA AAT TAT CAT GGA      240
141  Ser Trp Pro Leu Gln Asn Ala Asp Gly Thr Ala Ala Asn Tyr His Gly
142  65      70      75      80
143  TAT CAC ATC GTC TCC GCT TTA GCA GGT GAC CCA AAA AAC AGT GAT GAT      288

```

## RAW SEQUENCE LISTING

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:28

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

144	Tyr His Ile Val Ser Ala Leu Ala Gly Asp Pro Lys Asn Ser Asp Asp	
W--> 145	85 90 95	
146	ACT CCA CTT CAT TTA TTC TAT CAA AAA GTC GGT GAT ACA TCG ATT GAC	336
147	Thr Pro Leu His Leu Phe Tyr Gln Lys Val Gly Asp Thr Ser Ile Asp	
W--> 148	100 105 110 115	
149	AGC TGG AAA AAT GCT GGA AGA GTA TTT GAA GAT ATG GAT AAA TTT GTT	384
150	Ser Trp Lys Asn Ala Gly Arg Val Phe Glu Asp Met Asp Lys Phe Val	
W--> 151	120 125 130	
152	CCA AAT GAT CCG TAT CTT AAA TAT CAA ACA CAG GAG TGG TCA GGT TCT	432
153	Pro Asn Asp Pro Tyr Leu Lys Tyr Gln Thr Gln Glu Trp Ser Gly Ser	
W--> 154	135 140 145	
155	GCT ACT TTA ACC AAA GAT GGC CAA GTC CGT TTA TTC TAT ACA GAT TAC	480
156	Ala Thr Leu Thr Lys Asp Gly Gln Val Arg Leu Phe Tyr Thr Asp Tyr	
W--> 157	150 155 160	
158	TCA GGT AAT CCT GAA GAT GGT GGA ACC GGT GCT GGT AAC CAA ATC ATT	528
159	Ser Gly Asn Pro Glu Asp Gly Gly Thr Gly Ala Gly Asn Gln Ile Ile	
W--> 160	165 170 175	
161	TCA ACT GCT CAA GTA AAC TTA TCC CAG CCG GAT GCA GCT ACA CTT AAA	576
162	Ser Thr Ala Gln Val Asn Leu Ser Gln Pro Asp Ala Ala Thr Leu Lys	
W--> 163	180 185 190 195	
164	GTC GAT GGA GTA TCT GAT CAT AAA TCT GTC TTT GAT GGC GGA GAC GGT	624
165	Val Asp Gly Val Ser Asp His Lys Ser Val Phe Asp Gly Gly Asp Gly	
W--> 166	200 205 210	
167	ACA GTT TAT CAA AAT ATT CAG CAA TTT ATC GAT GAA GGC AAG TGG ATT	672
168	Thr Val Tyr Gln Asn Ile Gln Gln Phe Ile Asp Glu Gly Lys Trp Ile	
W--> 169	215 220 225	
170	TCA GGT GAT AAC CAT ACT TTA AGA GAC CCT CAC TAT GTT GAA GAT AAG	720
171	Ser Gly Asp Asn His Thr Leu Arg Asp Pro His Tyr Val Glu Asp Lys	
W--> 172	230 235 240	
173	GGC CAT AAA TAT CTT GTC TTT GAA GCG AAT ACT GGA ACA ACA GAT GGT	768
174	Gly His Lys Tyr Leu Val Phe Glu Ala Asn Thr Gly Thr Thr Asp Gly	
W--> 175	245 250 255	
176	TAT CAA GGC GAT CAG TCT TTC AAT AAT AAA GCT TAC TAT GGC GGA AGT	816
177	Tyr Gln Gly Asp Gln Ser Phe Asn Asn Lys Ala Tyr Tyr Gly Gly Ser	
W--> 178	260 265 270 275	
179	GAC GTC TTC TTC CAG AAT GAA AAA AAT AAA CTG CTT CAA AGT CCT AAA	864
180	Asp Val Phe Phe Gln Asn Glu Lys Asn Lys Leu Leu Gln Ser Pro Lys	
W--> 181	280 285 290	
182	AAA CAA ATT GCT TCT TTA GCG AAT GGT GCA TTA GGC ATT GTT GAA TTG	912
183	Lys Gln Ile Ala Ser Leu Ala Asn Gly Ala Leu Gly Ile Val Glu Leu	
W--> 184	295 300 305	
185	GCC GAT GAC TAT ACA GTG AAA AGT GTT ATG AAA CCA TTA GTC GCA TCA	960
186	Ala Asp Asp Tyr Thr Val Lys Ser Val Met Lys Pro Leu Val Ala Ser	
W--> 187	310 315 320	
188	AAC ACA GTA GCA GAT GAA GTC GAA CGC GCC AAT ATA TTT AAA ATG AAT	1008
189	Asn Thr Val Ala Asp Glu Val Glu Arg Ala Asn Ile Phe Lys Met Asn	
W--> 190	325 330 335	
191	AAT AAA TGG TAT CTA TTC ACG GAT TCA AGA GGA TCC AAA ATG ACG AGT	1056
192	Asn Lys Trp Tyr Leu Phe Thr Asp Ser Arg Gly Ser Lys Met Thr Ser	

## RAW SEQUENCE LISTING

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:28

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

```

W--> 193      340      345      350      355
      194      GAT GGA ATT AAC GAC AAA GAT GTT TAT ATG CTA GGG CCC GGA GGC GAC 1104
      195      Asp Gly Ile Asn Asp Lys Asp Val Tyr Met Leu Gly Pro Gly Gly Asp
W--> 196      360      365      370
      197      TCC TTA AAT GGC CCA CAC AAC CCG ATA AAT GAA ACT GGA CTT GTA TTG 1152
      198      Ser Leu Asn Gly Pro His Asn Pro Ile Asn Glu Thr Gly Leu Val Leu
W--> 199      375      380      385
      200      AAC ATG AAT CTT GAC CCT GCT GAT CTC ACA CAC ACT TAC TCT CAT TGC 1200
      201      Asn Met Asn Leu Asp Pro Ala Asp Leu Thr His Thr Tyr Ser His Cys
W--> 202      390      395      400
      203      GGT ATC CCG CAC CCT GAA GGT AAT AAT GTG GTA CTC ACA AGT TAT ATG 1248
      204      Gly Ile Pro His Pro Glu Gly Asn Asn Val Val Leu Thr Ser Tyr Met
W--> 205      405      410      415
      206      ACG AAT AGA GGC TTC TAT CCA GAA CAT CAC TCT CAC CTG CGG GAC AAG 1296
      207      Thr Asn Arg Gly Phe Tyr Pro Glu His His Ser His Leu Arg Asp Lys
W--> 208      420      425      430      435
      209      CTT GGG GTT AAT ATT AAA GGG TCT GAC ACA TCT GGA GGA GAA AAT AGT 1344
      210      Leu Gly Val Asn Ile Lys Gly Ser Asp Thr Ser Gly Gly Glu Asn Ser
W--> 211      440      445      450
      212      TCC GGA CAA GGA CAA TTC CCA 1365
      213      Ser Gly Gln Gly Gln Phe Pro
W--> 214      455
216 (2) INFORMATION FOR SEQ ID NO: 5:
217 (i) SEQUENCE CHARACTERISTICS:
218 (A) LENGTH: 2408 base pairs
219 (B) TYPE: nucleic acid
220 (C) STRANDEDNESS: double
221 (D) TOPOLOGY: linear
222 (ii) MOLECULE TYPE: genomic DNA
223 (vi) ORIGINAL SOURCE:
224 (A) ORGANISM: Bacillus sp.
225 (C) INDIVIDUAL ISOLATE: V230 (FERM BP-5054)
226 (ix) FEATURE:
227 (A) NAME/KEY: signal peptide
228 (B) LOCATION: 1..360
229 (C) IDENTIFICATION METHOD: E
230 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
231 CGGGGAAAAT ACTAGATTCC AATTGGGCCAG ACTTCCCAGT TGGTGTAAGA GAAGAGTTTCG 60
232 GACTGCCAAT GCAGCTGTGC GTAAGAAAAC AGCTTACTCA TGAGCAATTA CTAGAAGAAT 120
233 TTCAAAAAGTC CTGGGATAAG GCCAAGTCCA CTTTGAAATA AACTTTTCAG CCTCTGTGTG 180
234 GGGGCTTTTT TGTTTTATT TATTTCAACT GCAAGTGGTC CATCCCCTAT ATCAATTTAA 240
235 GACGAAATTC TAATCAATCC ATGCCATCCC CAATAAACTC GTCCTCCTCT ATACTTTTAA 300
236 TTAATAAGAA ACTATCAAGA GCTTCTTAT CAAATTCATA CATATCCAAG GAGGGAGACG 360
237 ATG AAC TTC AAA AGA TTG GCG AAA AAA GCA GCT GCC GTA ACC TTC AGG 408
238 Met Asn Phe Lys Arg Leu Ala Lys Lys Ala Ala Ala Val Thr Phe Arg
239 -30 -25 -20
240 ACT GCT ATA TTA GTA GGA GCG GAC GGA CCG CAT ATT TTT GCG CAG CAA 456
241 Thr Ala Ile Leu Val Gly Ala Asp Gly Pro His Ile Phe Ala Gln Gln
242 -15 -10 -5

```

## VERIFICATION SUMMARY

DATE: 02/25/2002

PATENT APPLICATION: US/09/986,682B

TIME: 08:41:29

Input Set : N:\Crf3\02122002\I986682A.raw

Output Set: N:\CRF3\02252002\I986682B.raw

L:20 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:21 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:145 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:148 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:151 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:154 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:157 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:160 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:163 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:166 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:169 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:172 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:175 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:178 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:181 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:184 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:187 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:190 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:193 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:196 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:199 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:202 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:205 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:208 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:211 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4  
L:214 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:4